DOCUMENT RESUME

ED 329 619 UD 027 812

AUTHOR	Quilling, Mary; Essl, Howard
TITLE	Using Evaluation Results for Local Reviews and Program Improvement.
INSTITUTION	Advanced Technology, Inc., Indianapolis, IN.; Chapter 1 Technical Assistance Center, Indianapolis, IN. Region B.
SPONS AGENCY	Office of Elementary and Secondary Education (ED), Washington, DC. Compensatory Education Programs.
REPORT NO	TAC-B-31
PUB DATE	90
NOTE	27p.
AVAILABLE FROM	Advanced Technology, Inc., 2601 Fortune Circle East, Indianapolis, IN 46241.
PUB TYPE	Guides - General (050)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Basic Skills; *Compensatory Education; *Educationally Disadvantaged; Elementary Secondary Education; *Evaluation Methods; Mathematics Skills; Measurement Objectives; *Outcomes of Education; Parent Participation; *Program Evaluation; *Program Improvement; Reading Skills; School Districts; Student Evaluation
IDENTIFIERS	*Education Consolidation Improvement Act Chapter 1

ABSTRACT

This guide focuses on the collection, measurement, and evaluation of data for programs operating under the mandate of Chapter 1 of the Education Consolidation and Improvement Act. Assessment of compensatory education performance of local education authorities (LEAs) is based on reporting of the following data: (1) basic and advanced skill scores; (2) demographic information; (3) student progress in regular programs; (4) continued progress in meeting desired outcomes; and (5) sustained effects. The guide lists possible additional data sources for program measurement and evaluation. A norm-referenced model is included. Tips on interpreting and using evaluation results and improving the accuracy of evaluation data are listed. Guidelines and sample reporting forms for a Chapter 1 annual review are included. A discussion of data disaggregation is provided. Data disaggregation is defined as the process of pulling apart test scores and other types of information concerning student achievement, instructional services, and staffing patterns in order to explore relationships between learning and program characteristics. This discussion includes guidelines for curriculum alignment and data disaggregation. A chart listing various scoring systems, a brief description of each system, and the advantages and limitations of each system is included. Checklists for Chapter 1 evaluation quality control and test selection are appended. (AF)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

CD
TAC-E-31
UD
Catalog

ED329619

1990
United States Department of Education
Chapter 1 Program Improvement Meetings

Using Evaluation Results for Local Reviews and Program Improvement

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Mary Quilling, TAC
Howard Essl, U.S. Department of Education/OPBE

Region B ESEA Chapter 1 Technical Assistance Center
Advanced Technology, Inc. • 2601 Fortune Circle East • Indianapolis, IN 46241
(800) 456-2380 • (317) 244-8160

UD027812

BEST COPY AVAILABLE

LEA EVALUATION

REPORTED TO SEA

1. Basic & Advanced Skill Scores
2. Demographic Information

MAINTAINED IN LEA (Results reported to SEA at least once every 3 years.)

3. Review Progress in Regular Program
4. Assess Continued Progress in Meeting Desired Outcomes
5. Sustained Effects

Basic and Advanced Skill Scores

1. Measure achievement in reading, mathematics, and language arts in grades 2-12

2. With regard to more advanced skills, assess progress of Chapter 1 participants as measured by the "comprehension" or equivalent score of a nationally normed reading test; and the "problems and applications" or equivalent score of a nationally normed mathematics test

3. In assessing achievement in language arts, use tests designed to measure language arts or reading. If a reading test is used, the LEA shall assess achievement in both basic and more advanced skills.

4. Exclude LEP projects designed to teach English to limited English speaking children

5. Measure student achievement over a period of approximately 12 months

6. Report on either a spring-to-spring or fall-to-fall testing interval

7. Aggregate by subject and grade for grades 2-12.

8. Report in the common reporting scale--NCEs

Demographic Information

1. Collect data on the race, age, gender, and number of children with handicapping conditions served by the program assisted under this chapter and on the number of children served by grade-level under the programs assisted under this chapter and annually submit such data to the Secretary. (P.L. 100-297, Sec. 1019.)

2. The SEA shall inform its LEAs, in advance, of the specific data that will be needed and how the data may be collected. (Federal Regulations, §200.35(b)(2).) An LEA shall provide to the SEA any data needed by the SEA to complete its annual performance report. (Federal Regulations, §200.35(c)(2).)

Progress in the Regular Program

Review may be based on--

- Teacher Judgments
- Grades
- Retention Rates
- Other Appropriate Indicators of Success

Desired Outcomes

1. Desired Outcomes are an LEA's goals to improve the educational opportunities of educationally deprived children to help those children--

   A. Succeed in the regular educational program
   B. Attain grade-level proficiency
   C. Improve achievement in basic and more advanced skills

2. At a minimum, must be expressed in terms of aggregate performance

3. May also use other indicators.

Sustained Effects

1. Collect information to determine whether student achievement gains are sustained over a period of more than 12 months.

2. Assess performance of the same children for at least two consecutive 12-month periods, provided these children continue to be enrolled in schools of the LEA.

3. Report information on either a spring-spring-spring or fall-fall-fall testing interval

POSSIBLE ADDITIONAL DATA SOURCES

Following are some possible sources of data to look at as you choose multiple measures for assessing Desired Outcomes, Progress in the Regular Program, Individual Student Progress, and Parental Involvement. Some sources will be appropriate for some purposes but not for others. Careful consideration should be given to the information each data source will contribute to your analysis.

ELEMENTARY PROGRAMS

- End of Unit/Text Tests
- Proficiency Tests
- State/District Criterion-Referenced Tests
- Teacher-Made Tests
- Performance on Classroom Assignments
- Number of Books Read
- Grade Placement--Success (or Lack of Success) in Being Promoted
- Performance on Homework Assignments
- Level in Basal Reader
- Mastery of Classroom Material
- Parent Involvement in Classroom Tutoring
- Parent Involvement in Home Learning Experiences
- Attendance at Parent Training Meetings
- Participation in Parent/Teacher Conferences
- Number of Parent Contacts With the School--letters, notes, phone calls

SECONDARY PROGRAMS

- End of Unit/Text Tests
- Proficiency Tests
- State/District Criterion-Referenced Tests
- Teacher-Made Tests
- Performance on Classroom Assignments
- Performance on Homework Assignments
- Mastery of Classroom Material
- Grade Placement--Success (or Lack of Success) in Being Promoted
- Grade Point Average
- Accrual of Credits Toward Graduation
- Graduation Rate
- Parent Involvement in Home Learning Experiences
- Attendance at Parent Training Meetings
- Participation in Parent/Teacher Conferences
- Number of Parent Contacts With the School--letters, notes, phone calls

MODEL A
NORM-REFERENCED MODEL

		OBSERVED		EXPECTED
TREATMENT	=	POST-TREATMENT PERFORMANCE	———	NO-TREATMENT PERFORMANCE

OBSERVED = TREATMENT GROUP'S MEAN STANDARDIZED SCORE ON NORMED POSTTEST

EXPECTED = TREATMENT GROUP'S POSTTEST EXPECTATION DETERMINED FROM PRETEST PERCENTILE

NORMAL CURVE EQUIVALENTS
WHY NCEs

To be useful for the purpose of measuring the impact of instructional treatments, a metric must be:

- Accurate in reflecting relative achievement levels
- Composed of equal-size units
- Sensitive to small gains
- Meaningful to the users

In addition, if comparisons are to be made or data to be aggregated, the metric must be able to combine results from different test instruments. NCEs were developed because the most widely used types of scores are deficient in one or more of these respects.

WHEN INTERPRETING EVALUATION RESULTS

- Use several comparisons; don't rely on just one standard or comparison in drawing conclusions for evaluative purposes.
- Calculate the median (not as sensitive to the distortion of small numbers and extreme scores) and the mean to evaluate gains.
- The accuracy of an NCE gain varies depending on the number of students on which the gain is based. Don't overinterpret gains based on small numbers.
- The size of NCE gain tends to be related to grade level (larger at lower grades) and subject area (larger for mathematics projects than reading projects).
- Differences in initial performance level may affect size of gain scores

IMPROVING ACCURACY OF EVALUATION DATA

- Track students to ensure the largest possible proportion of Chapter 1 students with both pre and posttests--less than 2/3 of students served is considered unrepresentative
- Test no more than 2 weeks from the empirical norming date; deviations from the norming date should be similar in direction and length for both pre and posttests
- Match the test used to your instructional program
- Choose a test sensitive enough to detect the effects of your project
- Administer the test in accordance with the publisher's instructions
- Ensure the best possible testing conditions by:
  - --eliminating distractions
  - --providing a comfortable setting
  - --providing a testing log for teachers to note any unusual occurrences
  - --familiarizing students with item format
- Examine carefully unusually high or low pretest and posttest NCE gains for possible sources of error and unreliability

PURPOSES OF EVALUATION

1. Detect individual strengths and weaknesses
2. Detect program weaknesses
3. Diagnose problems
4. PROGRAM IMPROVEMENT
5. Placement into programs
6. Determine extent of teaching success
7. Reporting to parents and others
8. Compete for awards/recognition

CHAPTER 1 ANNUAL REVIEW

Each school must conduct an annual review of data in 1 and 2, and possibly 3-6, and disseminate the information to parents, teachers, and others.

Must be discussed at annual review

School-level information

1. Aggregate gains for all educationally deprived students in grades 2 to 12, using Spring to Spring or Fall to Fall data.

    - Basic skills: record total reading and/or total math scores in NCEs
    - Advanced skills: record reading comprehension and/or mathematics problem solving and application scores in NCEs
    - Compute gains and losses and find the median or mean

2. Monitor and assess desired outcomes at all grade levels using the criteria established for "substantial progress." Set up databases by school for each outcome in your project application.

May be discussed at annual review

School-level information

3. Look at student level gains and attainment of desired outcomes and make program modifications for children who didn't gain.

District-level information

4. Look at performance in the regular program.
5. Conduct a sustained effects study every three years.
6. Assess parent involvement.

School-Level Annual Review

School: _____ **Lead Person:** _____

Review of pre-post test scores (aggregate performance for grades 2 and higher).

Instructional Area	NCE Gain*	Primary Focus Yes/No
Reading (Basic)		
Reading (Advanced)		
Math (Basic)		
Math (Advanced)		
Lang. Arts		

* Gain measured - spring to spring or fall to fall

Did the school reach the aggregate performance goals for the instructional area that is the primary focus of the school's Chapter 1 Program?

Yes ☐ No ☐

REVIEW OF DESIRED OUTCOMES STATED IN CHAPTER 1 APPLICATION

Desired Outcomes	% Attaining	Outcome Met? Yes/No
1.		
2.		
3.		
4.		
5.		
6.		

Did the school show "substantial progress" * toward meeting all desired outcomes? **Yes or No**

**Definition of "substantial progress" from the State Plan for Chapter 1 Improvement*

_____

_____

The school is identified to implement the Chapter 1 program improvement requirements (Sec. 200.38) if the school did not both reach the aggregate performance goals and make substantial progress toward achieving the stated desired outcomes.

School identified for program improvement? **Yes or No**

SUPPORT PAGE FOR AGGREGATE PERFORMANCE

	READING-BASIC			READING-ADVANCED			MATH-BASIC			MATH-ADVANCED			LANG. ARTS		
GRADE	N	NCE Ave. Gain	#Not Making Gain	N	NCE Ave. Gain	#Not Making Gain	N	NCE Ave. Gain	#Not Making Gain	N	NCE Ave. Gain	#Not Making Gain	N	NCE Ave. Gain	#Not Making Gain
2															
3															
4															
5															
6															
7															
8															
9															
10															
11															
12															
TOTAL															

Grade By Grade Support Page for Desired Outcomes

Desired Outcome # ___. _____

Grade	Attained Outcome		Did Not Attain Outcome		Total
	N	%	N	%	
K					
1					
2					
3					
.					
.					
12					
Total					

Was desired outcome attained?
Yes or No

Comment_____

_____

_____

_____

_____

_____

_____

_____

_____

Desired Outcome # ___. _____

Grade	Attained Outcome		Did Not Attain Outcome		Total
	N	%	N	%	
K					
1					
2					
3					
.					
.					
12					
Total					

Was desired outcome attained?
Yes or No

Comment_____

_____

_____

_____

_____

_____

_____

_____

_____

STUDENT PROGRAM IMPROVEMENT

1. Identify Students Whose Performance Shows no Improvement or Decline

2. Assess Their Performance Individually in Basic and Advanced Skills

   A. Must use standardized test results
   B. Use of multiple additional measures recommended

3. Conduct Thorough Assessment of Their Educational Needs and Use Results to Modify the Project to Serve Those Students Better.

LOCAL CONDITIONS

- The mobility of the student population
- The extent of educational deprivation among participating children that may negatively affect improvement efforts
- The difficulties involved in dealing with older children in Chapter 1 programs in secondary schools
- Whether indicators other than improved achievement demonstrate the positive effects on participating children of Chapter 1 activities
- Whether a change in the review cycle under section 1019 of the Act of paragraph (a)(1) of [§200.38] or in the measurement instrument used or other measure-related phenomena has rendered results invalid or unreliable for a particular year.

Student Identification

As part of the annual evaluation you are required to identify all students who have not shown substantial progress toward meeting desired outcomes or whose performance show no improvement or a decline.

1. List all students in the Chapter 1 program.
2. Indicate their current grade placement (during year when data was gathered).
3. Indicate a 1 for students not meeting goals for the first time, and a 2 if this is the second (or more) consecutive year in which the child was identified.
4. Place (X)'s to indicate areas in which the child did not improve.

Student Name	GR	Reading		Math		Lang Arts	Desired Outcomes						Will student be in program next year?		Is student identified for program improvement?		Number of years in program
		Basic	Adv.	Basic	Adv.		1	2	3	4	5	6	Yes	No	Yes	No	

INTERPRETING EVALUATION RESULTS

REQUIRED INDICATORS	STANDARDS
Did my program overall show satisfactory gains?	Aggregate project gains should equal standard set by state at a minimum
Did my project reach its goal?	Project objectives stated in measurable terms (i.e., desired outcomes)
How did my Chapter 1 students progress/succeed in the regular school program?	LEA standards based on teacher judgments, grades, retention rates, or other appropriate indicators

USEFUL INDICATORS	STANDARDS
How did my 3rd grade program do compared to my 3rd grade program last year?	Gains from last year's project at each grade
How did the 3rd grade program do compared to the 2nd and 4th grade Chapter 1 program?	Project gains at adjacent grade levels
How did my Chapter 1 3rd grade program compare to 3rd grade programs across the state?	Average gains for Chapter 1 projects across the state for the same grade and subject

USING EVALUATION RESULTS

ADMINISTRATORS DO YOU WANT TO . . .	LOOK AT . . .
1. identify classes which may need remediation?	summaries of subtest performance for each classroom.
2. find areas for curricular or instructional revision?	clusters of test objectives on which performance was low.
3. identify subject areas in need of revision or increased resources?	summaries of subject performance by grade for the school and/or district.
4. identify areas of improvement or decline?	subtest scores by grade on previous tests in the same area.
5. determine in-service needs?	subtests on which school or district performance was low.
6. compare achievement of your school and/or district with others?	narrative information by subtest and/or grade level
7. determine general achievement levels?	summaries of test performance with each grade by school.
8. know the general progress of students in your school and/or district?	NCE gains by grade level for the school and/or district.
9. compare student progress to the progress of others in the state or nation?	typical NCE gains made by other schools in the state or nation.
10. determine trends in student progress?	NCE gains compared to previous year.
11. find out which projects are most or least effective?	NCE gains for each project at each grade level.

DATA DISAGGREGATION[1]

Data disaggregation is the process of "pulling apart" test scores and other types of information concerning student achievement, instructional services, and staffing patterns in order to explore the relationships that exist between learning and program characteristics. It is one possible means of determining where to focus improvement efforts once a district has been identified for program improvement based on aggregate performance measures. Data disaggregation must be preceded by careful alignment of curriculum objectives with objectives covered by the norm-referenced test used to evaluate the Chapter 1 program. Following are steps to take in determining curriculum alignment and disaggregating data.

Curriculum Alignment

The process of determining curriculum alignment and subsequently using test scores to determine areas of weakness within the curriculum is a time-consuming process that requires the cooperation of staff across programs. Therefore, district administration must be willing to support staff members in this effort. In addition, the district curriculum must be articulated in some written form, such as school-based objectives, district-wide objectives (strategies or skill clusters), state syllabi, or some other type of defined format.

Curriculum Alignment is a four step process, as follows:

1. **Form a school-wide committee for each content area of interest.** Both the Chapter 1 and the regular school program should be represented on the committee, along with any school-wide or district curriculum and/or evaluation specialists.

2. **Develop two lists of objectives.** The committee should obtain lists of objectives or defined skills that are taught in each program and/or content area of interest. For purposes of curriculum alignment between the regular and Chapter 1 programs, the committee would need to obtain a list of objectives (strategies or skill clusters) for both programs in each content area being considered. If such lists do not exist, the committee will need to develop them. Each content area list should include 1) a summary statement for each objective, strategy, or skill, 2) some type of code number or reference number for each objective, strategy, or skill, and 3) an indication of approximately when each objective, strategy, or skill is taught in each program. This information is critical in the process of data disaggregation that follows.

---

[1] Based on Reichman, S. L. & Rayford, L., Using test results for curriculum alignment, an approach to program evaluation and improvement. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, April 5-9, 1988.

3. **Review the test and group related items.** A copy of the norm-referenced test needs to be obtained so that committee members can review actual test items and group related items by test objective (strategy or skill cluster). Any such groupings done by the test developer may be used as a basis, but the committee should review the groupings to ensure that it agrees with how the test publisher organized the strategies or skills. If the committee does not agree with the publisher's information, then it should make the requisite changes regarding the manner in which items are grouped.

   Each objective (strategy or skill cluster) assessed by the test should also be assigned a code number of a different but similar form from those assigned to curriculum objectives. The assessment objectivess are the link between the regular program and Chapter 1 instructional objectives, and therefore form the basis for reviewing curriculum alignment.

4. **Link test objectives to related program objectives.** Using the list of regular program and Chapter 1 objectives, along with the listing of test objectives and their respective code numbers, the committee should link each test objective to the related program objective by writing the appropriate test objective code number next to each program objective. This will likely require considerable committee discussion and perhaps some compromise. The results of this step indicate the congruence between objectives assessed on the test and the objectives taught in both the regular and Chapter 1 programs.

<u>Data Disaggregation</u>

The process of determining curriculum alignment as outlined above should be completed before meaningful disaggregation of data can occur. The process is as follows:

1. **Establish minimum passing level for the test and for each objective.** Based on the objectives (strategies or skill clusters) outlined by the curriculum alignment committee, establish a "minimum passing level" for the entire test and for each objective. The cut-off score should reflect the minimum standards that the committee feels students should attain in order to be classified as having passed the overall test and its individual objectives. The passing level for the test will be used as a basis for grouping students into mastery/non-mastery and should be based on total raw scores. Similar criterion levels should be set to determine whether individual students mastered or did not master each objective (strategy or skill cluster). Raw scores should be used to classify each student's success on each of the assessment objectives.

2. **Organize test results and disaggregate data.** Once the test has been administered, test results for the entire group should first be scored as follows: 1) **for each student** each item is scored as either correct or incorrect and the raw score indicating the total of items correct is obtained; 2) **for each student** a determination of passing/not passing is made for the

overall test and for each objective (strategy or skill cluster). There are various way of combining score information to obtain information about various aspects of your program. For instance you may want to look at student achievement overall and on each objective. You may want to examine building and classroom effects or effects of various service delivery models. You may consider dropout and graduation rates, program recidivism, behavior and discipline referrals, differences related to race, gender or ethnicity. You may look at effects of various instructional techniques on different objectives or strategies. For each question of this nature that you wish to examine, students' scores will be grouped differently and differences between groups examined.

3. **Interpret the results in view of the programs of interest.** Finally, the committee reviews the results of the disaggregation process to determine how well students are doing. As the committee reviews data, its task is to identify objectives (strategies or skill clusters) where there are performance discrepancies between groups of students; for instance, do students demonstrate varying levels of proficiency as a function of the type of educational services they receive? Or, does their test performance relate directly to the building(s) in which they receive those services?

   As is true with most test data used for program evaluation, the disaggregation process will highlight areas which should be investigated more closely, but will not provide specific information on potential solutions to problems that appear to exist. For example, results of the disaggregation process may indicate that a disproportionate number of students in the Chapter 1 program are not mastering certain reading comprehension strategies, even though the Chapter 1 objectives indicate that these strategies were taught to students prior to the testing date. A potential problem is highlighted by the disaggregation process, but the committee would have to delve further to determine why this is occurring.

CHARACTERISTICS OF VARIOUS SCORE SYSTEMS

TYPE OF SCORE	BRIEF DESCRIPTION	SOME ADVANTAGES	SOME LIMITATIONS
RAW SCORES	The number of correct responses obtained by the student. Range: 0-Total number of test items.	Could be used to assess mastery in a given area	Difficulty in comparing performance on subportions of the same tests Difficulty in comparing performance on different tests Difficulty in interpreting scores
PERCENTILES	The student's relative standing in the norm group in terms of the proportion of students scoring below him/her. Range: 1-99	Can be used to establish the standing of an individual relative to a normative group	Cannot be used in computation of group statistics. A difference of 10 percentiles at one point in the range is not equivalent to a different of 10 percentiles at another point on the range.
GRADE EQUIVALENTS	"Grade Placement" of student based upon test performance. Range: K-12	Apparent ease of communication (People believe they understand what is meant by Grade Equivalents)	Same as percentiles . plus: They are often misleading (e.g. a score of 7.0 by a fifth grade student does not mean he can do 7th grade work) Most G.E. values are estimates
STANDARD SCORES SYSTEMS	The amount by which a student's score departs from average. Range: depends on particular system (e.g. T-score; Z-scores; stanines, NCEs)	Performances of various tests and subtests can be computed. They are equal intervals.	Not all the publishers use standard scores or the same standard score system.
STANINES	A standard score system which has nine intervals	Same as standard score systems	Use of only nine intervals does not permit fine discrimination
EXPANDED SCALE SCORES	A standard score system with an arbitrary mid point but generally includes several grade levels. Range: Depends on particular system by test publisher	Same as standard score systems, plus: Comparison can be made from grade to grade Facilitates out of level testing	Not all publishers use them and not all publishers use the same numbering system. Can't combine results from different publisher's tests Difficulty of interpretation by lay-person
NORMAL CURVE EQUIVALENTS (NCEs)	A standard score system having 99 intervals. The average corresponds to the 50th percentile; the 1st and 99th NCEs correspond to the 1st and 99th percentiles	Same as standard score systems, plus: Ease of interpretation Permit aggregation of data from a wide variety of tests	They are relatively new They depend upon standard scores or percentiles Not all test publishers use them

QUALITY CONTROL CHECKLIST FOR CHAPTER 1 EVALUATION

Representativeness of Evaluation Findings

____ 1. All Chapter 1 participants, except those lost through attrition, are included in the evaluation pre- and post-testing.

____ 2. The evaluation plan includes procedures that should minimize the discrepancy between the number of persons served by the project and the number of persons for whom both pre- and post-test achievement scores are available.

- ____ a. A roster of participating Chapter 1 students is kept by the program evaluator or designee.
- ____ b. Procedures exist for notifying the program evaluator when Chapter 1 participants change classes, schools, etc., or move; reasons for the change; and the student's new location.
- ____ c. There is a plan for conducting make-up tests within the "acceptable" dates for the test(s) used.
- ____ d. The program evaluator is provided with reasons for missing data for each participant lacking pre- or post-test data.

Reliability and Validity of Evaluation Instruments and Procedures

____ 1. The instrument used for Chapter 1 evaluation has been reviewed in the last 3 years to ensure that it is appropriately matched to the curriculum of the program.

- ____ a. This process is documented.
- ____ b. The process was done by a committee including at least one Chapter 1 teacher, the Chapter 1 coordinator, a district administrator, and others as appropriate.
- ____ c. The instrument was chosen based on a match between objectives covered by the test and instructional objectives of the project.

____ 2. The edition of the test is either the current or immediate previous edition.

____ 2. Scoring of the test will be done in-house.

____ a. The answer sheets have been spot-checked to ensure that students filled them in correctly and that, if necessary, they can be accurately read by the scanner.

____ b. *Whether tests are hand- or machine-scored, someone* other that the person who initially did the scoring checks a sample of the tests for accuracy in scoring.

____ c. Conversion of raw scores to standard scores using conversion tables provided by the publisher is double checked for accuracy.

____ d. Any computer data entry is checked for errors and results from computer calculations are verified.

____ 3. Average NCE gains for each subject and grade are computed and verified for each building to ensure accuracy in data reported to the SEA.

____ 4. Aggregate NCE gains for each subject in each building are computed and verified for use in the Local Annual Review of program effectiveness.

Valid Interpretation and Use of Results

____ 1. Project staff understand the Chapter 1 metric (the NCE) used to measure Chapter 1 gains.

____ 2. Evaluation results are made available to project staff for review.

____ 3. Project staff use evaluation results to ensure the quality of testing and evaluation procedures AND to guide improvements in overall program quality.

____ 4. Project staff examine evaluation results for information about student needs at various grade levels and in different subject areas.

____ 4. Chapter 1 and regular classroom teachers review data diagnostically to plan instruction for Chapter 1 students.

____ 5. Parents and/or students have access to evaluation results and are provided with assistance in understanding them.

If out-of-level testing is necessary and machine-scoring is planned, does the scoring service provide a clear description of what information is required in order to obtain correct conversions?

- Will it be possible to obtain pretest and posttest scores for at least two-thirds of the students in the Chapter 1 program?

- Will appropriate test administration and scoring procedures be followed?

  Are the instructions clear for those (teachers/aides) who will administer the test?

  Is the length of time required to administer the test acceptable?

  Is the test group or individually administered, and which do you prefer?

  Will testing conditions be supportive of the students' best efforts during the pretest and the posttest?

  Will tests be scored accurately?

  Will average standard scores be calculated correctly only for those students with both pretest and posttest scores?

  Will average standard scores be converted correctly to percentiles and NCEs?

  Will the scoring procedure provide timely results?

CHECKLIST FOR CHOOSING A TEST FOR CHAPTER 1

- Does the content of the test match the content of the Chapter 1 program?

  Are most of the Chapter 1 program's instructional objectives measured by the test items?

  Are most of the test's objectives taught in the Chapter 1 program?

  Are the results reported in terms of the objectives?

- Is the test appropriate for the students?

  Will the students find the layout of the test booklet easy to follow?

  Is the reading level of the test appropriate?

  Will the students be able to understand the instructions?

  Is the response form simple enough?

- Does the test have representative, empirical norms?

  Does the norm group include students like the students in the Chapter 1 program?

  Does the norm group include school systems likely to have the regular instructional programs like your school system in the appropriate grades?

  Have you examined the most recent or immediately preceding edition of the test?

- Does the test have empirical norming dates that are within two weeks of when the pretest and posttest are to be administered?

- Will at least two-thirds of the program occur between pretesting and posttesting?

- Does the test have a level which matches the functional level of the students in your Chapter 1 program?

  What is the functional level of your students?

  If out-of-level testing is necessary and hand-scoring is planned, does the person responsible for obtaining average standard scores and percentiles understand the proper conversion procedures?